Reliability and Validity of Different Models of TKK Hand Dynamometers

Cristina Cadenas-Sanchez, Guillermo Sanchez-Delgado, Borja Martinez-Tellez, José Mora-Gonzalez, Marie Löf, Vanesa España-Romero, Jonatan R. Ruiz, Francisco B. Ortega

OBJECTIVE. We examined the reliability and validity of the analog and digital models of TKK handgrip dynamometers using calibrated known weights.

METHOD. A total of 6 dynamometers (3 digital and 3 analog; 2 new and 1 old for each model) were used in this study.

RESULTS. Intrainstrument reliability was very high; systematic error for test–retest reliability was ≤0.3 kg. The systematic error among different instruments (same model) and between different models (digital vs. analog) ranged between |0.4 kg| and |0.6 kg|. The systematic error between new and old dynamometers ranged from |0.8 kg| to |1 kg|. All dynamometers provided lower values for the same known weights than a SECA scale, with a systematic error ranging from −2.04 kg to −2.64 kg.

CONCLUSION. This study indicates that clinicians and investigators who provide treatment to address handgrip strength should use the same instrument and model for repeated measures. Distinguishing meaningful change from dynamometer variability is discussed.

Muscular strength is considered an important marker of health (Blair et al., 1989; Ortega, Ruiz, Castillo, & Sjöström, 2008; Ruiz et al., 2009). Growing evidence indicates that muscular strength is inversely associated with muscular disorders, back pain, osteoarthritis, and premature mortality (Bergman et al., 2001; Ortega, Silventoinen, Tynelius, & Rasmussen, 2012; Timpka, Petersson, Zhou, & Englund, 2013). Szeto and Lam (2007) examined whether musculoskeletal discomfort was associated with strength factors and concluded that a weaker left handgrip was strongly associated with neck and shoulder pain. Furthermore, growing evidence indicates that physical workload and other occupational factors are a risk factor for many musculoskeletal disorders and discomfort in several parts of the body (Rachiwong, Panasiriwong, Saoomphop, Widjaja, & Ajjimaporn, 2015; Roquelaure et al., 2009; Szeto & Lam, 2007). Therefore, the assessment of muscular strength as a health indicator has become important from a clinical, therapeutic, occupational, and public health perspective.

The handgrip strength test, which is the most reliable and valid field-based muscular fitness test (Artero et al., 2011; Castro-Pinero et al., 2010), has traditionally been assessed using the Jamar dynamometer and has been recommended by the American Society of Hand Therapists (Shechtman & Sindhu, 2013). Some studies have focused on the reliability and validity of the Jamar dynamometer (Gerodimos, 2012; Härkönen, Harju, & Alaranta, 1993; King,
Our research group has previously determined the reliability and validity of the Jamar, DynEx, and digital TKK dynamometers using calibrated known weights (España-Romero et al., 2010). We observed that the digital TKK showed the highest test–retest reliability and the highest criterion-related validity (i.e., smaller mean difference when compared with known weights). Differences between TKK and Jamar dynamometers showed that the TKK dynamometer is more reliable and valid than the Jamar (España-Romero et al., 2010); the grip span of the TKK dynamometer can be continuously adjusted to differences in hand size using age- and gender-specific equations (España-Romero et al., 2008; Ruiz et al., 2006; Ruiz-Ruiz, Mesa, Gutiérrez, & Castillo, 2002; Sanchez-Delgado et al., 2015), whereas the Jamar has five positions (beside the position of the hand; see Supplemental Figure 1, available online at http://otjournal.net; navigate to this article, and click on “Supplemental”). The TKK dynamometer does not need regular calibration, but the Jamar requires calibration every year.

On the basis of this evidence, our recommendation is to use the digital TKK dynamometer. However, the digital version has a range that measures from 5 to 100 kg, which might not be useful in populations whose handgrip strength values can be lower than 5 kg, such as people with certain pathologies or work-related hand injuries, people in hand therapy programs, older people, or preschool children. For these clients, the analog version of the TKK dynamometer, which measures from 0 to 100 kg, would be a good alternative. To the best of our knowledge, the reliability and validity of the analog TKK have not been examined objectively (i.e., measured with calibrated known weights). We identified several relevant questions regarding the reliability and validity of hand dynamometers that still need to be addressed—in particular, how reliable the handgrip measure is when using two different instruments of the same model, different models, or a new dynamometer versus an older one. This information is important for researchers, clinicians, physiotherapists, and sport scientists because this measure provides important information about client prognosis (Savino et al., 2013) and may be pivotal in task-oriented rehabilitation programs for chronic patients (da Silva, Antunes, Graef, Cechetti, & Pagnussat, 2015).

The overall aim of the current study was to examine the reliability and validity of the digital and analog models of TKK dynamometers using calibrated known weights. The specific aims of the study were to determine test–retest reliability within instruments of the same TKK model (i.e., intrument reliability), interinstrument reliability between instruments within the same TKK model (e.g., between two digital dynamometers and 2 analog dynamometers), intermodel reliability (digital vs. analog dynamometers), reliability of new TKK dynamometers versus old ones, and validity of dynamometers against calibrated known weights.

Method

Instruments

A total of 6 TKK handgrip dynamometers, 3 digital and 3 analog, were used to assess the reliability and criterion-related validity. The digital handgrip dynamometers (TKK Model 5401; Takei, Tokyo, Japan) had a range of measure from 5.0 kg to 100.0 kg, whereas the analog handgrip dynamometers (TKK Model 5001; Takei, Tokyo, Japan) had a range of measure from 0.0 kg to 100.0 kg (Supplemental Figure 2, available online at http://otjournal.net; navigate to this article, and click on “Supplemental”).

Four dynamometers were new (i.e., bought for this study), and 2 (1 digital and 1 analog) were old (i.e., had been used for >6 yr in population studies), allowing us to compare the reliability and validity between new and old TKK dynamometers. The dynamometers, weights, and scale were calibrated by the manufacturer at purchase. The verification of all weights was performed by means of a new (bought for this study) high-precision SECA scale (Model 769; SECA, Hamburg, Germany). We assumed that the SECA scale was perfectly calibrated because we could not test its validity against a gold standard. However, we assessed its reliability using known weights from 1 to 70 kg with increments of 1 kg up to 20 kg and increments of 5 kg thereafter. We observed very high reliability; that is, the mean difference between the known weights and the SECA scale measures was 0.004 kg (standard deviation [SD] = 0.02 kg; \(p = .300\)).

Procedures

Known weights were used to analyze the criterion-related validity (known weights vs. dynamometers) and reliability

---

**Table 1.** The overall mean difference between the known weights (\(\Delta\)) and the dynamometers (SECA scale measurements).

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Mean</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamar</td>
<td>0.04</td>
<td>0.02</td>
<td>.300</td>
</tr>
<tr>
<td>TKK</td>
<td>0.04</td>
<td>0.02</td>
<td>.300</td>
</tr>
</tbody>
</table>
(intrainstrument, interinstrument, and intermodel measures) of the 6 dynamometers (Supplemental Figure 3, available online at http://otjournal.net; navigate to this article, and click on “Supplemental”). The dynamometers were positioned between two wooden supports with the handle fixed. The known weights were suspended with a loading belt from the center of the dynamometer’s handle; the weights ranged from 1 kg to 70 kg, and increments of 1 kg were added up to 20 kg and increments of 5 kg thereafter. The dynamometer’s handle was marked for consistent placement of the loading belt with the known weight. The weights were added in a randomized order, and each weight measure was repeated twice (test–retest). The order of testing of the dynamometers was also randomized. As commonly described in the literature, a 5.0-cm grip span was used (roughly corresponding to Position 3 of a Jamar grip span). The time between trials was approximately 50–60 s.

Statistical Analysis

The agreement among intrainstrument, interinstrument, and intermodel trials (i.e., reliability) and the agreement between the known weights and dynamometer measures (i.e., criterion-related validity) were assessed using Bland and Altman’s (1986) method. Mean difference (error) and the 95% limits of agreement (error ± 1.96 of the difference) were calculated. Results were graphically examined by plotting the differences against their mean (Bland & Altman, 1986). One-sample t test was used to test whether the mean difference (i.e., systematic error) was significantly different from zero (reference).

On the basis of results from previous studies (Bénèfice, Fouére, & Malina, 1999; Molenaar, Zuidam, Selles, Stam, & Hovius, 2008) and a study conducted by our group in preschool children (ages 3–5 yr; Sanchez-Delgado et al., 2015), we observed that preschoolers’ handgrip strength is likely to be less than 15 kg. To gather information about how reliable and valid the dynamometers would be in people with hand injuries, older adults, and preschoolers, we used one-way analysis of variance (ANOVA; with intertrial mean difference as the dependent variable) to test whether the intertrial mean differences of the dynamometers studied were significantly different at weights ≤15 kg versus >15 kg.

Heteroscedasticity, or whether the variability (error) increases or decreases as the magnitude of the measures changes, is an important dimension of an instrument’s reliability and validity. To calculate the heteroscedasticity, the negative values of the difference (both reliability and validity analyses) were changed to positive (i.e., multiplied by −1) and fitted into a one-way ANOVA model, with weight groups as the fixed factor (i.e., light weights ≤15 kg and heavier weights >15 kg). A significant difference (p < .05) between light and heavier weights would confirm heteroscedasticity.

An exploratory analysis was performed to know whether the selection of the grip span influenced the reliability of the dynamometers. We measured known weights (5, 10, 20, 30, 40, 50, 60, and 70 kg) at several grip spans (4.0, 4.5, 5.0, and 5.5 cm). The measures were performed once for each model of dynamometer (new digital and new analog). For all the analyses, the level of significance was set at p < .05.

Results

Reliability

Table 1 presents the mean differences among repeated measures with the same instrument (intrainstrument reliability), different instruments of the same model (interinstrument reliability), different models (i.e., digital vs. analog; intermodel reliability), and old versus new dynamometers. Regarding intrument reliability, the mean difference ranged from 0.04 kg to 0.25 kg for the digital dynamometer and from 0.09 kg to 0.33 kg for the analog dynamometer. The systematic error between different instruments within the same model (e.g., digital 2 vs. digital 1; interinstrument reliability) was 0.6 kg for both the digital and analog dynamometers. Systematic error ranged from 0.24 kg to 0.35 kg for the comparison between different new models (i.e., digital vs. analog). The systematic error between old and new digital dynamometers was 1.08 kg for the digital and 0.78 kg for the analog dynamometers. All reliability estimates did not differ between weights ≤15 and >15 kg (p > .05).

We tested for heteroscedasticity for all the comparisons shown in Table 1 and observed that the intrainstrument variability between measures increased as the magnitude increased for the new analog 1 (≤15 kg = 0.14 ± 0.21; >15 kg = 0.60 ± 0.65; p = .014) and new analog 2 (≤15 kg = 0.14 ± 0.29; >15 kg = 0.49 ± 0.56; p = .041) dynamometers. No significant heteroscedasticity was observed for the rest of the dynamometers. Reliability analyses are graphically shown using Bland–Altman plots in Figures 1 and 2. Plots are shown for one dynamometer only per model; similar plots were obtained for the other dynamometers (data not shown). In the exploratory analyses, we observed that the differences among grip spans (from 4.0 to 5.5 cm) ranged from 0.00 kg to 0.66 kg and from 0.00 kg to 0.75 kg for the new digital and new analog dynamometers, respectively (Supplemental Table 1,
Table 1. Differences in Reliability Among TKK Dynamometers

<table>
<thead>
<tr>
<th>Dynamometer</th>
<th>General</th>
<th>Difference Between Weights, $p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M \pm SD$</td>
<td>weights $\leq 15$ kg, $M \pm SD$</td>
</tr>
<tr>
<td>Comparison using the same instrument (intrainstrument reliability: retest minus test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New digital 1</td>
<td>0.04 ± 0.60</td>
<td>0.04 ± 0.58</td>
</tr>
<tr>
<td>New digital 2</td>
<td>−0.06 ± 0.55</td>
<td>−0.16 ± 0.26</td>
</tr>
<tr>
<td>Old digital</td>
<td>−0.25 ± 0.65</td>
<td>−0.30 ± 0.48</td>
</tr>
<tr>
<td>New analog 1</td>
<td>0.09 ± 0.65</td>
<td>0.06 ± 0.25</td>
</tr>
<tr>
<td>New analog 2</td>
<td>−0.22 ± 0.53</td>
<td>−0.13 ± 0.30</td>
</tr>
<tr>
<td>Old analog</td>
<td>−0.33 ± 0.69</td>
<td>0.08 ± 0.48</td>
</tr>
<tr>
<td>Comparison between different instruments of the same model (interinstrument reliability)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New digital 2 minus new digital 1</td>
<td>−0.82 ± 2.03</td>
<td>−0.54 ± 1.13</td>
</tr>
<tr>
<td>New analog 2 minus new analog 1</td>
<td>−0.64 ± 1.33</td>
<td>−0.23 ± 0.64</td>
</tr>
<tr>
<td>Comparison between different models (intermodel reliability)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New digital 1 minus new analog 1</td>
<td>−0.35 ± 1.51</td>
<td>−0.01 ± 0.92</td>
</tr>
<tr>
<td>New digital 2 minus new analog 2</td>
<td>−0.25 ± 1.72</td>
<td>−0.31 ± 0.63</td>
</tr>
<tr>
<td>Comparison of old vs. new dynamometers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old digital minus new digital 1</td>
<td>1.08 ± 2.02</td>
<td>1.00 ± 1.44</td>
</tr>
<tr>
<td>Old analog minus new analog 1</td>
<td>−0.78 ± 1.32</td>
<td>−1.13 ± 0.84</td>
</tr>
</tbody>
</table>

Note. $M =$ mean; $SD =$ standard deviation.

*a*One-sample $t$ test. The intertrial difference was entered as a dependent variable; $p$ value indicates whether the mean difference is significantly different from 0 for all measures from 1 kg to 70 kg. $^b$One-way analysis of variance. The intertrial difference was entered as a dependent variable, and weights ≤15 kg vs. >15 kg as a fixed factor; $p$ value indicates whether the mean difference is the same or different for weights ≤15 kg (i.e., people with lower handgrip strength, such as adults with work-related hand injuries or preschoolers) compared with weights >15 kg (i.e., people with higher handgrip strength, such as healthy children age >6 yr, adolescents, and adults).

available online at http://otjournal.net; navigate to this article, and click on “Supplemental”).

**Validity**

Criterion-related validity for the dynamometers against known weights is provided in Table 2. A negative systematic error (underestimation) was observed for the new digital (range = −2.64 kg to −2.02 kg) and new analog (range = −2.15 kg to −1.51 kg) dynamometers as well as for the old digital and analog dynamometers (−0.94 kg and −2.29 kg, respectively). The systematic error increases as the magnitude of the measures increases (heteroscedasticity) in all of the dynamometers except the old digital one. Most had a mean difference range from 0.97 to 1.83 kg for weights ≤15 kg and from 2.22 to 3.62 kg for weights >15 kg.

**Discussion**

The current study provides several major findings. First, the intrainstrument test–retest reliability was excellent for all of the dynamometers. The combined systematic error of ≤0.3 kg is relatively low for digital and analog devices, especially for the analog dynamometer, which has a precision of measure of 0.5 kg. Second, the systematic error between different instruments of the same model and between different models (digital vs. analog) ranged between 0.3 to 0.6 kg, with higher errors of measure observed between old and new dynamometers (both digital and analog). This finding is in accordance with the finding by Härkönen et al. (1993) that older Jamar dynamometers were less accurate than newer ones. This fact raises a point of caution when interpreting results of handgrip strength levels in long-term follow-up studies because slight improvements (roughly 1 kg) or impairments in hand therapy or exercise programs could be attributable to a systematic error between new and old dynamometers. Thus, these findings should be interpreted and generalized cautiously because the systematic error might differ depending on the frequency of use, the way the dynamometer is used, and how old the dynamometer is.

Third, the systematic error between new and old dynamometers ranged from 10.8 kg to 11.1 kg, which in our opinion is a relatively large error. Finally, all dynamometers provided lower values than known weights.

The main message drawn from these findings is that whenever possible, clinicians should use the same dynamometer in repeated measures to minimize the systematic error inherent in the apparatus. If repeated measures are taken using different instruments (e.g., new digital 2 minus new digital 1) or different models (e.g., new digital minus new analog), the systematic error is expected to range from 0.3 to 0.6 kg, with higher errors of measure observed between old and new dynamometers (both digital and analog). This finding is in accordance with the finding by Härkönen et al. (1993) that older Jamar dynamometers were less accurate than newer ones. This fact raises a point of caution when interpreting results of handgrip strength levels in long-term follow-up studies because slight improvements (roughly 1 kg) or impairments in hand therapy or exercise programs could be attributable to a systematic error between new and old dynamometers. Thus, these findings should be interpreted and generalized cautiously because the systematic error might differ depending on the frequency of use, the way the dynamometer is used, and how old the dynamometer is.

The heteroscedasticity analysis showed significant differences in the new analog dynamometers between light and heavier weights, which indicates that the test–retest
error increases as the client’s handgrip strength increases. A practical implication of this finding is that when assessing handgrip strength using the analog TKK dynamometer in a specific population (e.g., people with certain pathologies or hand injuries, preschoolers) with very low handgrip strength, the systematic error will be very small (i.e., <0.2 kg). This error is smaller than that of the digital version (i.e., 0.5 kg), which supports the use of analog TKK dynamometers in populations with very low handgrip strength, with the additional advantage that the analog dynamometer measures less than 5 kg, whereas the digital one does not. In contrast, when assessing older children and adolescents or healthy and young adults, the digital version would be a better choice because of its high reliability and lack of heteroscedasticity. Our exploratory analyses indicate that the same grip span should be used when testing the same person repeatedly because mean differences among grip spans might be up to 0.8 kg.

The criterion-related validity analyses suggest that all dynamometers studied provide lower values than the SECA scale (−0.94 to −2.64 kg), which could be interpreted as a slight underestimation. This result is in contrast with what we observed in our previous study, in which the TKK (only the digital model) overestimated handgrip strength levels (0.49 kg; España-Romero et al., 2010). The difference between studies could be partially attributed to the model of the weight scale used (SECA 769 vs. 861) or to interinstrument error (described in this study). Nevertheless, we believe this last finding is not very relevant because handgrip strength will never be measured using a body...
weight scale. In addition, and in line with previous research (Mullaney, Darmstadt, Khatry, Leclerq, & Tielsch, 2007), we assumed that the SECA scale was the gold standard, and therefore we calibrated the weights using this scale; however, we could not know how valid the SECA scale is because no better gold standard was available to test it. Consequently, we can conclude only that the TKK dynamometer provides values 1–2 kg lower than the SECA scale. Overall, heteroscedasticity analysis showed significant differences for the digital and analog dynamometers indicating that the error of the measures increased as the magnitude of the measures increased, as previously described (España-Romero et al., 2010).

Our results strongly support the use of the same instrument in repeated measures to minimize the systematic error to 0.3 kg or less. If different instruments, models, or dynamometers of different ages are used in repeated measures, the systematic error might increase up to 1 kg. This information is useful for studies in which the aim is to compare handgrip strength levels over different time periods (e.g., before and after a hand therapy program, in intervention or follow-up studies). For example, in the AVENA and HELENA studies, Moliner-Urdiales et al. (2010) analyzed secular trends in adolescents’ health-related physical fitness between 2001 and 2007. Adolescents showed a change of 4 kg in strength values measured by the handgrip strength test. If the researchers used the same dynamometer 6 yr later, the expected error would be 1 kg. Some additional variance would be attributable to the biological variability of each participant, but the 4 kg of secular change reported is likely to reflect a real change in strength.
The main strength of the current study is that we used TKK dynamometers, which have been shown to be reliable and valid and also to be sufficiently precise to detect minimum changes in comparison with other dynamometers (e.g., Jamar). A limitation of this study is the absence of a group of people with different characteristics to examine the reliability of different models of dynamometers. Further studies involving human samples and different types of dynamometers (e.g., Jamar, DynEx) are needed to confirm our findings.

Implications for Occupational Therapy Practice

The results of this study have the following implications for occupational therapy practice:

- The TKK dynamometer is a useful tool for hand grip assessment with good reliability and validity and higher precision, reliability, and validity than other dynamometers such as the Jamar (España-Romero et al., 2010).
- Practitioners providing hand therapy should use the analog version of TKK, rather than the digital version, because it allows assessment of hand grip strength in patients with very low strength levels.
- Findings support the recommendation to use the same instrument to measure hand strength because interinstrument reliability adds a certain amount of error.
- This study provides objective estimates of systematic error so that practitioners, researchers, and anyone who needs to compare data from two different time points can distinguish between the variability of the instrument itself and a meaningful change in hand grip strength attributable to an intervention program or hand therapy.

Acknowledgments

The study took place under the umbrella of the Active-Brains project funded by the Spanish Ministry of Economy and Competitiveness (Reference DEP2013–47540). We thank Christine Delisle for her help with the English proofreading. Cristina Cadenas-Sanchez is supported by a grant from the Spanish Ministry of Economy and Competitiveness (BES–2014–068829). Jonatan R. Ruiz and Francisco B. Ortega are supported by grants from the Spanish Ministry of Science and Innovation (RYC–2011–09011 and RYC–2010–05957, respectively). In addition, the costs related to this project were covered with support from the Spanish Ministry of Science and Innovation (RYC–2011–09011). This work is part of a PhD thesis in biomedicine at the University of Granada, Granada, Spain. All authors declare that they have no conflicts of interest to disclose.

References


---

Table 2. Differences Between TKK Dynamometers and Known Weights

<table>
<thead>
<tr>
<th>General</th>
<th>Weights ≤15 kg, M ± SD</th>
<th>Weights &gt;15 kg, M ± SD</th>
<th>Difference Between Weights, p²</th>
</tr>
</thead>
<tbody>
<tr>
<td>New digital 1</td>
<td>-2.02 ± 2.60 .001</td>
<td>-0.73 ± 1.03</td>
<td>-2.89 ± 2.99 .010</td>
</tr>
<tr>
<td>New digital 2</td>
<td>-2.64 ± 2.73 .001</td>
<td>-1.27 ± 0.90</td>
<td>-3.65 ± 3.17 .003</td>
</tr>
<tr>
<td>Old digital</td>
<td>-0.94 ± 3.05 .135</td>
<td>0.27 ± 1.29</td>
<td>-1.75 ± 3.62 .105</td>
</tr>
<tr>
<td>New analog 1</td>
<td>-1.51 ± 1.73 .001</td>
<td>-0.71 ± 0.27</td>
<td>-2.31 ± 2.18 .009</td>
</tr>
<tr>
<td>New analog 2</td>
<td>-2.15 ± 2.20 .001</td>
<td>-0.93 ± 0.60</td>
<td>-3.37 ± 2.54 .001</td>
</tr>
<tr>
<td>Old analog</td>
<td>-2.29 ± 1.27 .001</td>
<td>-1.83 ± 0.83</td>
<td>-3.14 ± 1.48 .047</td>
</tr>
</tbody>
</table>

Note. M = mean; SD = standard deviation.

*One-sample t test. The intertrial difference was entered as a dependent variable; p value indicates whether the mean difference is significantly different from 0 for all measures from 1 kg to 70 kg. *One-way analysis of variance. The intertrial difference was entered as a dependent variable, and weights ≤15 kg versus >15 kg as a fixed factor; p value indicates whether the mean difference is the same or different for weights ≤15 kg (i.e., people with lower hand grip strength, such as adults with work-related hand injuries or preschoolers) compared with weights >15 kg (i.e., people with higher hand grip strength levels, such as healthy children age >6 yr, adolescents, and adults).


